Exam Code(s)   3IF1

Exam(s)   3rd B.Sc. in Information Technology

Module Code(s)   CT332

Module(s)   Database Systems II

Paper No.
Repeat Paper

External Examiner(s)   Prof. Michael O' Boyle
Internal Examiner(s)   Dr. J. Duggan
                                  Mr. C. O' Riordan

Instructions   Answer any 3 questions
                                 All questions carry equal marks.

Duration   3 hours
No. of Answer books   1

No. of Pages   4
Department(s)   Information Technology

Q.1.

i)   Describe the process of normalisation. Decompose the following relation R, given the functional dependencies in F, into relations such that all relations satisfy BCNF.

R = {A, B, C, D, E, F, G, H, I, J}

F ={ {ABC} → {D, E},
      {E} → {C},
      {AB} → {F},
      {C} → {G},
      {F} → {H},
      {H} → {I, J},
      {F} → {B} }   (9)

ii)   Explain what is meant by the term *lossless join property*. Consider the following relation:

EMP_DEPT {ssn, emp_name, emp_dept_no, salary, dname, dlocation}

and the functional dependencies:

{{ssn} → {emp_name, emp_dept_no, salary},
  {emp_dept_no} → {dname, dlocation}}

suggest a suitable decompostion of the relation and show that the decomposition has the lossless join property.   (8)

iii)   Outline an algorithm for generating a minimal cover set from a set of functional dependencies. Illustrate the operation of your algorithm with the following set of functional dependencies.

F = {AB → D, B → C, AE→ B, A→ D, D→ E, D→ F}

                                                              (8)

iv)   Design by synthesis is one approach to designing a database given the universal relation R and a set of functional dependencies. Conceptual design followed by mapping to a suitable logical design represents an alternative approach to designing a relational database. Discuss the limitations and strengths of these two approaches.   (8)

Q.2.

i)   Timestamping and two-phase locking are two approaches to ensuring concurrency
      control. Outline *either* of the two approaches and present pseudo-code for the primitives
      used.
      Show how the following schedule would proceed under *either* protocol.

Ta	**Tb**	**Tc**
		read_item(X)
		read_item(W)
read_item(X)		
read_item(Y)		
read_item(Z)		
	read_item(W)	
	write_item(W)	
	read_item(Z)	
	write_item(Z)	
write_item(Y)		
write_item(X)		
write_item(Z)		
		write_item(W)

                                                                    (16)

ii)   Define the term *conflict-serializability.* Show that enforcing timestamp ordering guarantees
      that schedules will be conflict-serializable.   (9)

iii) Outline an approach to guaranteeing the atomic nature of transactions in distributed
      databases. Show how this approach operates with different types of possible failures.
                                                                    (8)

Q.3.

i)   Given the following database schema:

                USER: user_id, fname, sname, address, age
                RATES: user_id, rest_id, date, score
                REVIEWS: user_id, rest_id, date, review
                RESTAURANT: rest_id, name, address, type, openingtime, closingtime

        Develop an SQL query for the following information need:

        List all restaurants that have been reviewed by "Daniel Johnston" *or* have both a type "Italian" and at least one rating score of at least 4.

        Develop an operator tree that represents an efficient evaluation strategy for the above query.
                                        (11)

ii)   Discuss the structure of a B-tree and describe an algorithm for insertion of values into a B tree. Illustrate the algorithm by showing how a tree (with order 3) would develop given the following numbers to be inserted:

                9, 10, 11, 5, 3, 24, 23, 29

                                              (11)

iii)   Outline an efficient algorithm for implementation of a join operator. How can the join operator be implemented more efficiently given a parallel database?   (11)

Q.4.

i)   Illustrate with examples, how the relational operators (select, project, join) may be implemented in Datalog. Discuss, briefly, the differences between the expressiveness of SQL and the expressiveness of Datalog.   (12)

ii)   With reference to the database schema presented in Q.3, provide SQL code to generate a view entitled **GOOD_RESTAURANTS** which would list all restaurants with at least 5 reviews and an average rating score of at least 4.
        Explain the concept *view update problem* and illustrate how it occurs with the view **GOOD _RESTAURANTS**.   (11)

iii)   With respect to mandatory access control in databases, explain what is meant by the *simple security property* and the *star property*. Illustrate with an example how violation of these principles can lead to security problems in the database.   (10)